CLAIMS

1. A method of generating a definition for a collection of source documents comprising:

identifying patterns common to each source document in the collection of source documents; and

constructing for an element type in the collection of source documents a restrictive general rule based on the identified common patterns.

2. The method of claim 1, wherein identifying common patterns comprises:

identifying common attribute names and types.

3. The method of claim 2, wherein identifying common patterns further comprises:

identifying restricted attribute values associated with the common attribute names and types.

4. The method of claim 2, wherein identifying common attribute names and types comprises:

determining the number of occurrences of each attribute name on an element type;

examining the attribute values for each occurrence of each attribute name on the same element type to determine the attribute type; and

determining if the attribute name occurs in association with the same attribute value on more than one element type.

5. The method of claim 3, wherein identifying restricted attribute values comprises:

examining attribute values for each occurrence of an attribute type in all of the source documents in the collection of source documents; and

establishing an enumeration or a restricted range appropriate to the attribute type.

6. The method of claim 5, wherein identifying restricted attribute values further comprises:

applying a heuristic to identify errors in the collection of source documents; and

adjusting the established enumeration or restricted range for attribute values.

7. The method of claim 1, wherein constructing a restricted general rule comprises:

constructing a content model that specifies the sequence order and number of occurrences of sub-elements within the common pattern.

8. The method of claim 2, wherein constructing a restricted general rule comprises:

constructing attribute definitions and value rules for each identified common attribute name and type.

9. The method of claim 1, further comprising:

identifying those patterns found to achieve a predetermined threshold of commonness; and

constructing a restrictive general rule for those identified patterns.

10. A computer program residing on a computer-readable medium for building a document type definition for a collection of source documents, the computer program comprising instructions causing a computer system to:

identify patterns common to each source document in the collection of source documents; and

construct for an element type in the collection of source documents a restrictive general rule based on the identified common patterns.

11. The computer program of claim 10, wherein the instructions to identify common patterns comprise instructions to:

identify common attribute names and types.

12. The computer program of claim 11, wherein the instructions to identify common patterns further comprise instructions to:

identify restricted attribute values associated with the common attribute names and types.

13. A computer system comprising:

a storage device for storing a set of source documents; and

a computer processor configured by a document type definition building program to identify patterns common to each source document in the set of source documents and construct for an element type in the set of source documents a restrictive general rule base on the identified common patterns.

14. A method of converting a format of a first source document to a format of a similarly structured second source document, the method comprising:

identifying patterns common to the first and second source documents; and

using the identified common patterns to map elements and sub-elements in the first source document to equivalent elements and sub-elements in the second source document.

15. The method of claim 14, further comprising:

replacing tag names for each of the elements and sub-elements in the first source document with equivalent tag names of the elements and sub-elements in the second source document.

16. The method of claim 14, wherein identifying patterns common to the first and second source documents comprises:

examining document type definitions for the first and second source documents.

17. The method of claim 16, further comprising:

producing the document type definition for the first source document if the document type definition for the first source document does not already exist.

18. The method of claim 14, wherein identifying patterns common to the first and second source documents comprises:

performing pattern matching.

19. The method of claim 14, wherein identifying patterns common to the first and second source documents comprises:

matching heuristics of the patterns in the first source document to heuristics of the patterns in the second source document.

20. The method of claim 18, wherein identifying patterns common to the first and second source documents further comprises:

matching heuristics of the patterns in the first source document to heuristics of the patterns in the second source document.

21. The method of claim 14, wherein using uses the identified common patterns to map automatically elements and sub-elements in the first source document to equivalent elements and sub-elements in the second source document.

22. A method of converting the format of a source document

to the format of a set of source documents, the set of source documents having a structure similar to the first source document, the method comprising:

identifying patterns common to the source document and the set of source documents;

mapping elements and sub-elements in the common pattern of the source document to equivalent elements and sub-elements the common pattern of the set of source documents; and

replacing tag names for the each of the elements and sub-elements in common pattern of the source document with the equivalent tag names of the elements and sub-elements in common pattern of the set of source documents.

23. The method of claim 22, wherein identifying patterns common to the source document and the set of source documents comprises:

examining document type definitions for the source document and and the set of source documents.

24. The method of claim 23, further comprising:

producing the document type definition for the source document if the document type definition for the source document does not already exist.

25. A computer program residing on a computer-readable medium for converting a format of a first source document to a format of a similarly structured second source document, the computer program comprising instructions causing a computer

system to:

identify patterns common to the first and second source documents; and

use the identified common patterns to map elements and sub-elements of the first source document to equivalent elements and sub-elements of the second source document.

26. The computer program of claim 25, further comprising instructions to:

replace tag names for the each of the elements and sub-elements in the common pattern of the first source document with equivalent tag names of the elements and sub-elements in the common pattern of the second source document.

27. The computer program of claim 26, wherein the instructions to identify patterns common to the source document and the set of source documents comprise instructions to:

examine document type definitions for the source document and and the set of source documents.

28. A computer system comprising:

a storage device for storing a source document and a set of source documents, the source document having a format different from that of the set of source documents; and

a computer processor configured by a mapping program to identify patterns common to the source document and the set of source documents and map elements and sub-elements in the common pattern of the source document to equivalent elements and sub-

9 elements the common pattern of the set of source documents.